Accurate Automated Segmentation of Autophagic Bodies in Yeast Vacuoles Using Cellpose 2.0

Emily C. Marron[a], Jonathan Backues[b], Andrew M. Ross[a*], and Steven K. Backues[c*]

[a] Department of Mathematic and Statistics, Eastern Michigan University, Ypsilanti, MI, USA

[b] Euler Scientific, 5452 Sonoma Place, San Diego, CA, 92130, USA

[c] Department of Chemistry, Eastern Michigan University, Ypsilanti, MI, USA

*To whom correspondence should be addressed: aross15@emich.edu or sbackues@emich.edu

Abstract

Segmenting autophagic bodies in yeast TEM images is a key technique for measuring changes in autophagosome size and number in order to better understand macroautophagy/autophagy. Manual segmentation of these images can be very time consuming, particularly because hundreds of images are needed for accurate measurements. Here we describe a validated Cellpose 2.0 model that can segment these images with accuracy comparable to that of human experts. This model can be used for fully automated segmentation, eliminating the need for manual body outlining, or for model-assisted segmentation, which allows human oversight but is still five times as fast as the current manual method. The model is specific to segmentation of autophagic bodies in yeast TEM images, but researchers working in other systems can use a similar process to generate their own Cellpose 2.0 models to attempt automated segmentations. Our model and instructions for its use are presented here for the autophagy community.

Keywords

automated labeling; autophagy; computer vision; electron microscopy; image analysis; machine learning

Abbreviations

AB, autophagic body; AvP, average precision; GUI, graphical user interface; IoU, intersection over union; MVB, multivesicular body; ROI, region of interest; TEM, transmission electron microscopy; WT, wild type

This is an original manuscript of an article published by Taylor & Francis in Autophagy on 18 May 2024, available online at https://doi.org/10.1080/15548627.2024.2353458

Introduction

When an autophagosome fuses with the yeast vacuole, its inner membrane and contents are delivered into the vacuole lumen, forming an autophagic body (AB). Normally, ABs are rapidly degraded by vacuolar hydrolases. However, the inhibition of protease activity in the vacuole, for example by genetic removal of the activating protease Pep4, allows ABs to accumulate so they can be measured [1]. The size and number of ABs in a vacuole corresponds to the size and number of autophagosomes that were formed. Therefore, measuring ABs in different conditions or different mutants can provide valuable insights into the process of autophagosome formation [2-6].

The only microscopy technique with a high enough resolution to allow measurement of AB size is transmission electron microscopy (TEM). However, because regular TEM visualizes only a thin slice of the cell, many hundreds of cells must be imaged to provide enough data for accurate estimation of original body size and number [7]. Manually labeling the ABs in these images is a laborious process that takes many hours of expert time, creating a clear need for an automated solution.

Rapid advances in computer vision over the past decade have led to the hope that an algorithm could be developed to automatically recognize and label ABs in TEM images. However, this has proven to be a particularly difficult task. Structures in TEM images, unlike those in many fluorescent images, are identified by subtle differences in contrast and texture. Regular TEM images show only a slice of the cell, so there are no three-dimensional data to aid in identifying complete structures. (Collecting 3D TEM images through techniques such as serial sectioning is possible but challenging, and not practical for the large sample sizes we need to accurately assess average AB size and number.) Finally, ABs can have other cellular membranes inside them that can look like the boundary of a body, while at the same time imperfect TEM preservation can make the actual boundaries incomplete. Together these factors can make it challenging to distinguish one body from the next, and even experts vary somewhat in their labeling of a given image. Despite these challenges, we found that a custom model created using the cellular segmentation program Cellpose 2.0 was able to label autophagic bodies in TEM images with an accuracy comparable to that of human experts.

Results

The application of computer vision to scientific images is an area of active research, and over the past years we screened a number of tools to see if any were suitable to this particular task. Our screening criteria were threefold. First, the tool should be targeted at our problem (object recognition in TEM images), or something similar and easily generalizable. Secondly, it should have good documentation and be easy to use, with the features we need. Key features included the potential to work in a fully automated fashion, the ability for the user to easily correct the automated output, and a simple method for exporting the results for downstream measurement. Third, the tool should give promising initial results with minimal training to show that it is on the right track and worth the effort of further optimization.

Prior to 2022, the available tools we tested did not meet these criteria. Cellprofiler [8], a widely used program that can achieve excellent results with fluorescent microscopy images by

application of methods such as threshold and watershed, was not able to segment ABs in our tests, likely because the characteristics of TEM images are quite different. Similarly, the MorpholibJ library of tools, as well as other tools available in ImageJ, were not able to approach this problem in our hands [9,10]. In general, it seemed that traditional methods of automated image analysis were not up to the task, and that a machine learning approach was likely to be necessary. WEKA is a trainable machine-learning tool that integrates into ImageJ [11]. However, it also gave poor initial results on our images and did not seem suitable to this problem. Slightly more promising was Ilastik, a suite of tools specifically designed for segmenting TEM images [12]. Ilastik's trainable Multicut tool could occasionally recognize bodies accurately, but it was optimized for segmenting an entire image and therefore gave many spurious results, had somewhat limited user control, and did not seem to improve with training on our images. We also tested a proprietary commercial solution from a computer vision company, which created a custom model trained on 100 of our images. It performed vastly better than the other tools, likely due to this more extensive training, but was still quite a bit worse than a human segmenter and ultimately not good enough to be useful in practice. Stardist, a publicly available machine learning tool, was optimized for detecting whole cells in fluorescence microscopy images and did not give promising results on our TEM images [13]. Segment Anything, in contrast, is a completely generalist machine learning image segmentation model recently made publicly available [14]. Unfortunately, it also gave poor initial results, likely because it is too general and not sufficiently optimized for our images, which are more challenging than average. Of course, these tools are under continual development, and it is possible that with new updates some of them may soon (or even now) be suitable to our task. However, the first program that we were able to find that met our screening criteria with promising initial results was Cellpose 2.0 [15].

Cellpose 1.0, published in 2021, was a generalist segmentation tool that consisted of a neural network model pretrained on a large and diverse set of microscopy images combined with a GUI (graphical user interface) that allowed easy user editing and exporting of the results [16]. It introduced a new watershed method based on vector flows that helped it find the center and edges of structures of a predefined size, and the model they trained outperformed Stardist on the images that they tested. Although described as a generalist model, it was primarily trained and tested on fluorescent microscopy images and did not prove suitable for segmenting autophagic bodies in TEM images. Cellpose 2.0, published in 2022, extended the original Cellpose tool with a key capability: it allowed users to train their own specialized models for segmenting specific types of images according to their needs [15]. Even though Cellpose 2.0 was not specifically designed for labeling TEM images, we decided to try training it on our TEM images of ABs.

Traditional machine learning methods require training on a large number of images to reach optimal performance. Cellpose 2.0 bootstraps this process with the use of partially pretrained models; the user selects a model from the provided model zoo and then trains it further with their own labeled images. Our preliminary tests with the pretrained models in the model zoo indicated that "CPx" with a diameter parameter of 130 pixels (where our images are 2.16 nm per pixel) gave the most promising initial results. This is not surprising given that CPx was the model pre-trained on the widest variety of images, while the other models were trained on exclusively fluorescent or differential interference contrast images [15]. The initial results with CPx were already better than those achieved with any other tools we tried, and the Cellpose program had all of the features that we needed. At the same time, the results with CPx alone were clearly not good enough for actual use, so more training on our specific images was needed.

In previous reports, Cellpose 2.0 was able to achieve optimal training on fluorescent microscopy segmentation problems using only 500-1000 labeled regions of interest (ROIs) when beginning with a pretrained model then proceeding to further train on specific images [15]. In those applications, an ROI is typically a cell or nucleus, while in our context an ROI is a profile of an AB in a thin TEM section. The images we segment typically contain between one and 40 AB profiles per image, averaging around 10, raising the possibility that as few as 50-100 images might be sufficient for training. To determine if this estimate would apply to recognizing ABs in electron microscopy images, we trained four separate models using different numbers of labeled ABs. We also trained each model for different numbers of epochs (1 to 999) to determine how many were required for optimal performance; using too many training epochs runs the risk of overtraining the model, which can reduce its performance on novel images. After training, each model was tested on a set of 51 images, and the model's automated labels were compared to the human-labeled ground truth using the same calculation that Cellpose uses: the average precision (AvP) at intersection over union (IoU) thresholds of 0.5 and 0.75 [15]. (Note that previous publications abbreviated "average precision" as AP, but here we use AvP to avoid confusion with "autophagosome" for readers of this journal.) An IoU of 0.5 means that 50% of the total pixels of both labels of the same object were in agreement with each other and indicates that the human and machine had identified roughly the same structure, while an IoU of 0.75 is a "good" agreement between two labels of the same structure. Therefore, if a model had an AvP score of 0.65 at IoU threshold = 0.5 and 0.45 at IoU threshold = 0.75, that means it labeled 45% of the ABs the same as the reference segmentation and labeled a total of 65% fairly similarly.

Figure 1A shows how the accuracy of the models changed during training. We found that each model's accuracy reached a plateau by about 100 epochs, and evidence of overtraining started around 300 epochs. In addition, we found that increasing the number of ROIs did not significantly enhance the model's performance: Model 2, trained on 1400 ABs, performed only slightly better than Model 0, trained on 350; Model 3, trained on 1837 ABs, actually performed slightly worse than Model 2 (Figure 1A). This is consistent with previous claims that Cellpose 2.0 requires relatively few ROIs for training and is beneficial should other researchers desire to train their own models. We selected Model 2 at 201 epochs as our best performing model for further testing and use.

Model 2 had an AvP score of 0.65 at IoU threshold = 0.5. This is significantly better than the original CPx model, which had an AvP score of only 0.124 at IoU threshold = 0.5 on our images, underscoring the importance of the training we performed. However, published Cellpose 2.0 models for fluorescent or differential interference contrast images were able to reach AvP scores of 0.7 to 0.8 at IoU threshold = 0.5 [15]. One explanation for our lower score (0.65) would be if TEM images of ABs were particularly difficult to segment even for humans. To test this, we measured between-human agreement on AB labeling by asking eight human experts to segment nine challenging test images. We found the average human AvP score to be only 0.49 at IoU threshold = 0.5 and 0.35 at IoU threshold = 0.75 (Figure 1B). This underscores the fact that segmentation of ABs in these images is somewhat ambiguous, and individual scientists may interpret them differently. We used Model 2 to segment those same nine test images, and it achieved an AvP score of 0.53 at IoU threshold = 0.5 and AvP of 0.40 at IoU threshold = 0.75, suggesting that it can perform this segmentation on par with human experts (Figure 1B,C).

Even though Cellpose 2.0 performs as well as a typical expert, a given researcher might want to revise Cellpose's segmentations to make sure they conform to their own best judgment.

Fortunately, this is also easy to do, thanks to Cellpose 2.0's GUI that allows rapid editing of the segmentation. To estimate how much time could be saved by this sort of "model-assisted segmentation," we used Cellpose 2.0's GUI with Model 2 loaded to segment 20 new test images one at a time, manually editing any segmentations we were not completely satisfied with. That took a total of 12 min of computation time plus 7 min of human time for the editing. In contrast, manually segmenting the same 20 images took 39 min. Therefore, model-assisted segmentation took less than one fifth as much human time as fully manual segmentation – a major speed improvement.

ABs will continue to accumulate in the vacuole over the course of an extended starvation. At times, vacuole cross-sections are found that are completely full of AB cross-sections. Our training and test images were from a 3-h nitrogen starvation, and only ~6% of these showed cross-sections that were mostly or entirely full. To specifically test if Cellpose 2.0 would still be able to accurately recognize ABs after longer starvations or when they fill the vacuole cross-section, we tested Model 2 on two additional sets of test images. The first set was 40 random images from cells that had undergone 4 h of nitrogen starvation (~15% of these showed cross-sections that were mostly or entirely full). The second set was 20 images from either 3 or 4 h of nitrogen starvation that had been handpicked because they showed vacuole cross-sections that were mostly or entirely full of AB cross-sections. In both cases, Model 2 performed well - not quite as well as it had on the original 3-h starvation images, but as well as it had on the nine challenging test images chosen to test the segmentations of various experts (Figure 2A-C). Model 2's slightly poorer performance on these new images is not surprising, because sets of images from separate experiments may differ slightly in appearance and preservation. Overall, this suggests that longer starvation times and higher number of bodies is not a significant challenge to our Cellpose model, even when a vacuole section is completely full of autophagic body sections.

One of the main factors that Cellpose models consider when identifying objects is the size (in pixels) of the desired features. ABs come in a range of sizes, and some mutants affect average AB size [2,4]. This raises the question of whether our model might struggle to recognize larger or smaller ABs. To partially address this, we included images in our training dataset from a mutant with reduced expression of Atg7 that gives ~20% smaller average AB cross-sections than wild type (WT) cells [6]. When Model 2 was tested on only these images, its performance was equivalent to its performance when tested on only WT type images (Figure 2D), demonstrating that this small size difference was not an issue. We then analyzed the range of sizes of AB cross-sections identified by Cellpose in all test images, and found that 99% of the cross-sections it labeled were between 60 and 600 nm in diameter, with the remaining 1% being larger, up to 850 nm (Figure 2E). ABs formed by bulk macroautophagy are around 400-900 nm in diameter [17], while some of the smallest bodies formed by selective macroautophagy (cytoplasm-to-vacuole targeting/Cvt pathway vesicles) are 150 nm [18]. Of course, random cross-sections taken from these structures are mostly smaller than those maximum diameters. This suggests that Model 2 should be able to recognize cross-sections from most ABs formed by either bulk or selective macroautophagy without difficulty.

In addition to the ABs formed via bulk and selective macroautophagy, starvation induced microautophagy - which involves the invagination of the vacuolar membrane instead of the *de novo* formation of an autophagosome - also gives rise to ABs. These microautophagic bodies, of which there are different types, can range in size from around 25 nm to 500 nm in diameter [19].

While the larger end of this range falls into that recognized by Model 2, the lower end does not, suggesting that this model may not be appropriate for recognizing ABs resulting from microautophagy. Similarly, multivesicular bodies (MVBs) are ~25 nm in diameter [19], and would likely not be recognized. It should be noted that formation of both microautophagic bodies and MVBs is dependent on the ESCRT machinery, including Vps4. Our training and test images were taken from strain lacking Vps4 (this is commonly done in order to eliminate the background from these processes when studying macroautophagy). Therefore, Model 2 is trained only to recognize ABs from macroautophagy, not those from microautophagy or MVBs.

In order to determine whether Model 2 would also recognize microautophagic bodies and MVBs, despite not being trained to do so, we tested it on 90 images of starved cells that contained Vps4 but were deleted for *ATG9*. Each of these images contain many small lumenal vesicles that likely represent MVBs and/or microautophagic bodies, but should contain no macroautophagic bodies due to the lack of Atg9. Often these vesicles were found in clusters within the vacuole (Figure 2F). The vast majority of these vesicles and clusters were not recognized by model 2; only 16 total vesicle clusters were labeled by our Cellpose model out of the 90 total images. In addition, the model mistakenly identified 22 other structures in these 90 images, 10 of which resemble ABs, 3 of which are other structures in the vacuole, and 9 of which were structures not found in the vacuole. For researchers using our model to study macroautophagy, this represents a relatively small background of misidentified structures, some of which might also fool a human. Our model would not be suitable for researchers more interested in microautophagy or the MVB pathway, as it is not good at recognizing those structures. However, our experience suggests that a new Cellpose 2.0 model could likely be trained to recognize such structures of interest.

The only limitation that we found to Cellpose 2.0's performance on our task is that it was not very good at recognizing images that had no ABs in the vacuole (important for estimation of AB number). It was able to correctly identify only about 70% of empty vacuoles, placing spurious ABs in the others, and this performance did not improve at all with training (Figure 2G). This is likely because we could not include any images with 0 AB in the training data, as the neural network is not set up to train on 0 AB/ROI. It is beyond the scope of this work to make any major changes to the Cellpose 2.0 code. Fortunately, this limitation does not significantly impede the usefulness of Cellpose 2.0 for segmentation of ABs because empty vacuoles can be quickly and easily recognized by humans and require no segmentation, so they can simply be separated from the images to segment before application of Cellpose.

Discussion

Manual segmentation of large numbers of TEM images has been a rate-limiting factor in the analysis of how different mutations affect autophagosome size and number, as is true of many TEM-based techniques. This not only slows the pace of research but also limits the practical sample size, potentially reducing the accuracy of the final results. In addition, individual variations in segmentation style may lead to issues when comparing results between researchers. Fortunately, advances in machine learning have now made it possible for a computer algorithm to segment ABs in TEM images with reasonable accuracy. The Cellpose 2.0 model that we present here is as accurate as an independent human expert, but much faster and with reproducible results.

In addition, this model can be further customized as needed. The Cellpose 2.0 GUI allows a user to easily edit the automatically generated segmentations. Moreover, those edited segmentations can be used to further train the model to adjust its style, a procedure described by the Cellpose authors as "human-in-the-loop" training [15]. Because the model requires only a few hundred ROI's to reach essentially maximum performance, it is even feasible to train an entirely new model if desired. However, using the model presented here without further training or editing of the results would have the advantage of allowing greater reproducibility between studies.

Our recommendation is to first try the provided model on a few of the desired images. If results are satisfactory, it can then be used as is to automatically label all images. If the user is dissatisfied with some of the labels generated, these can be easily altered using the Cellpose 2.0 GUI, and the user can proceed with model-assisted segmentation, or, if desired, train their own model instead.

How to measure ABs using Cellpose 2.0 for labeling

1. Acquire TEM images of AB in yeast vacuoles, carefully following the steps described in section 2.1 of Backues et al. 2014 [7]. To avoid potential biases, the images should be labeled using a code so that the downstream analysis can be done in a blinded fashion.
2. Manually sort out images of vacuoles that do not contain any bodies. These will still be used for the analysis of AB number, but do not need to be segmented.
3. Download the trained AB model from https://osf.io/wrez4/
4. Download, install and run Cellpose 2.0 to segment the images containing ABs
    a. For simplest use, install the Cellpose 2.0 GUI using instructions from https://github.com/MouseLand/cellpose. The Cellpose 2.0 GUI allows images to be segmented one at a time: Load the AB model, load the desired image, choose "run model" to generate masks, alter these if desired, and then save them as png/tif.
    b. For those with more computational experience, automated segmentation can be performed faster in batch using Pycharm and a Docker container (biocontainers/cellpose:2.1.1_cv2); utilities for this can be found at https://github.com/StevenBackues/cellpose_AB.
5. Copy the mask files to their own folder and use these to measure the area of each labeled AB and the number of ABs per vacuole. This can be easily done in python using code provided in the cellpose_AB utilities in GitHub and in Google Colab at https://colab.research.google.com/drive/1y1Z5k7bKxJ3PsExf_e5IYFMKMxF3aDpS. Alternatively, it can be done in ImageJ using the MorphoLibJ library [9,10].
6. Estimate the original size and number distribution of the bodies from this data following the instructions in section 2.3 of Backues et al. 2014 [7].

Methods

Images of ABs used for the training and test sets were obtained and segmented as previously described [7]. In brief, yeast were starved for 3 or 4 h in medium without nitrogen (SD-N) to induce autophagy, and prepared for TEM via chemical fixation. Each image was captured at a magnification of 30,000x, yielding images of 2240 x 2240 pixels, with a scale of 1 px = 2.16 nm. Each image included one yeast near the center, and some had small portions of other yeast near the edges or corners, but no image showed two entire yeast vacuoles. Ground truth manual

segmentations were performed by J., S. or R. Backues and verified by S. Backues before use. Details on the strains and images used for training and testing are included in the Table 1 and Table 2, while the images themselves, including ground truth segmentations, have been made publicly available at https://osf.io/tuhwn/.

For stand-alone use via the GUI, Cellpose 2.0 was downloaded from its official Github repository, https://github.com/mouseland/cellpose. For batch labeling, model training and testing, we used Pycharm Professional and Docker Desktop with the Cellpose 2.0 Docker container biocontainers/cellpose:2.1.1_cv2. All training and tests were run using an nVidia RTX 4090 24GB graphics card; the model can also be used on other systems (GPU or CPU) with essentially identical results. Details of code used can be found at https://github.com/StevenBackues/cellpose_AB.

We thus proceeded to train "CPx" into custom models using different sized sets of AB training data (350, 700, 1400 and 1837 ROIs) and training for 1 to 1000 epochs. Stratification was used to ensure a diverse number of ABs per image (but always more than 0) in both the test and training sets. In addition, we used an empty-vacuole image test set which contained 33 images with no AB. Average Precision (AvP) at different IoU thresholds was measured using Cellpose's built-in testing utility [16], with AvP scores averaged over all test images. The proportion of empty vacuoles correctly identified was calculated using a simple Boolean where True corresponds to correctly predicting that an image had no AB, and False corresponds to any failure to do so (i.e. predicting the presence of any AB in the empty vacuoles). The number of correct predictions was divided by the total number of empty vacuoles tested (33) to give the final score.

Table 1. Strains used in this study.

Name	**Strain genotype**	**Source**
atg9Δ	SEY6210 *pho13Δ pho8Δ60 atg9Δ::LEU2 pep4Δ::KAN*	Jin et al. 2014 [5]
pho23Δ	SEY6210 *vps4Δ::TRP1 pep4Δ::LEU2 pho23Δ::KAN*	Jin et al. 2014 [5]
Reduced Atg7 (E1)	SEY6210 *vps4Δ::TRP1 pep4Δ::LEU2 atg7Δ::HIS5 ura3-52::pRS406-GAL3p-ATG7-PA*	Cawthon et al. 2018 [6]
rim11Δ	SEY6210 *vps4Δ::TRP1 pep4Δ::LEU2 rim11Δ::URA3*	This study
SEY6210	MATα *his3Δ200 leu2-3,112 lys2-801 suc2-Δ9 trp1Δ901 ura3-52*	Robinson et al. 1998 [20]
WT	SEY6210 *vps4Δ::TRP1 pep4Δ::LEU2*	Cheong et al. 2005 [2]
Wild type (B4)	SEY6210 *vps4Δ::TRP1 pep4Δ::LEU2 atg1Δ::HIS5 ura3-52::pRS406-ATG23p-ATG1-PA*	This study
Wild type (B6+C2)	SEY6210 *vps4Δ::TRP1 pep4Δ::LEU2 atg7Δ::HIS5 ura3-52::pRS406-ATG7p-ATG7-PA*	Cawthon et al. 2018 [6]

An unpublished analysis from our lab has shown the size of AB cross-sections in cells expressing Atg1-PA under the *ATG23* promoter ("Wild Type B4") to be indistinguishable from cells expressing Atg1-PA under the native *ATG1* promoter, thus we classify this strain as wild type in regards to AB size.

Table 2. Images used in this study.

Image Set	Strain name	Starvation time	# of images
Model 2 (training)	Wild type (B4)	3 h	48
	Wild type (B6+C2)	3 h	75
	Reduced Atg7 (E1)	3 h	9
3 h starvation (initial testing)	Wild type (B4)	3 h	16
	Wild type (B6+C2)	3 h	31
	Reduced Atg7 (E1)	3 h	4
Wild type (testing)	Wild type (B6+C2)	3 h	31
Reduced Atg7 (testing)	Reduced Atg7 (E1)	3 h	8
4 h starvation (testing)	*rim11Δ*	4 h	20
	WT	4 h	18
Filled vacuoles (testing)	*pho23Δ*	3 h	5
	rim11Δ	4 h	7
	WT	4 h	8
atg9Δ (testing)	*atg9Δ*	3 h	90
Empty vacuoles (testing)	Wild type (B4)	3 h	21
	Wild type (B6+C2)	3 h	11
	Reduced Atg7 (E1)	3 h	1

Acknowledgements: Many thanks to Hayley Cawthon, Ronith Chakraborty, Elizabeth Delorme-Axford, Payton Dunning, Yuchen Feng, Jacquelyn Roberts, Patrick Wall and Zhiping Xie for serving as the human experts for comparing AB segmentation variability. Thanks to Dr. Dan Klionsky for providing some of the raw TEM images used for extended model testing. Thanks to Rebecca Backues for performing manual segmentation of some of the images used for training, and to Mark Backues for helpful discussions in the early stages of this project.

Works Cited

1. Takeshige K, Baba M, Tsuboi S, Noda T, Ohsumi Y. Autophagy in yeast demonstrated with proteinase-deficient mutants and conditions for its induction. J Cell Biol 1992; 119:301–11.

2. Cheong H, Yorimitsu T, Reggiori F, Legakis JE, Wang C-W, Klionsky DJ. Atg17 Regulates the Magnitude of the Autophagic Response. Mol Biol Cell 2005; 16:3438–53.

3. Xie Z, Nair U, Geng J, Szefler MB, Rothman ED, Klionsky DJ. Indirect estimation of the area density of Atg8 on the phagophore. Autophagy 2009; 5:217–20.

4. Xie Z, Nair U, Klionsky DJ. Atg8 controls phagophore expansion during autophagosome formation. Mol Biol Cell 2008; 19:3290–8.

5. Jin M, He D, Backues SK, Freeberg MA, Liu X, Kim JK, Klionsky DJ. Transcriptional regulation by Pho23 modulates the frequency of autophagosome formation. Curr Biol CB 2014; 24:1314–22.

6. Cawthon H, Chakraborty R, Roberts JR, Backues SK. Control of autophagosome size and number by Atg7. Biochem Biophys Res Commun 2018; 503:651–6.

7. Backues SK, Chen D, Ruan J, Xie Z, Klionsky DJ. Estimating the size and number of autophagic bodies by electron microscopy. Autophagy 2014; 10:155–64.

8. Kamentsky L, Jones TR, Fraser A, Bray M-A, Logan DJ, Madden KL, Ljosa V, Rueden C, Eliceiri KW, Carpenter AE. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. Bioinformatics 2011; 27:1179–80.

9. Legland D, Arganda-Carreras I, Andrey P. MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. Bioinformatics 2016; 32:3532–4.

10. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012; 9:671–5.

11. Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, Sebastian Seung H. Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics 2017; 33:2424–6.

12. Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, Schiegg M, Ales J, Beier T, Rudy M, et al. ilastik: interactive machine learning for (bio)image analysis. Nat Methods 2019; 16:1226–32.

13. Schmidt U, Weigert M, Broaddus C, Myers G. Cell Detection with Star-convex Polygons [Internet]. 2018 [cited 2024 Feb 21]. page 265–73.Available from: http://arxiv.org/abs/1806.03535

14. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, et al. Segment Anything [Internet]. 2023 [cited 2024 Feb 21]; Available from: http://arxiv.org/abs/2304.02643

15. Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. Nat Methods 2022; 19:1634–41.

16. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. Nat Methods 2021; 18:100–6.

17. Baba M, Takeshige K, Baba N, Ohsumi Y. Ultrastructural analysis of the autophagic process in yeast: detection of autophagosomes and their characterization. J Cell Biol 1994; 124:903–13.

18. Baba M, Osumi M, Scott SV, Klionsky DJ, Ohsumi Y. Two Distinct Pathways for Targeting Proteins from the Cytoplasm to the Vacuole/Lysosome. J Cell Biol 1997; 139:1687–95.

19. Schuck S. Microautophagy – distinct molecular mechanisms handle cargoes of many sizes. J Cell Sci 2020; 133:jcs246322.

20. Robinson JS, Klionsky DJ, Banta LM, Emr SD. Protein sorting in Saccharomyces cerevisiae: isolation of mutants defective in the delivery and processing of multiple vacuolar hydrolases. Mol Cell Biol 1988; 8:4936–48.

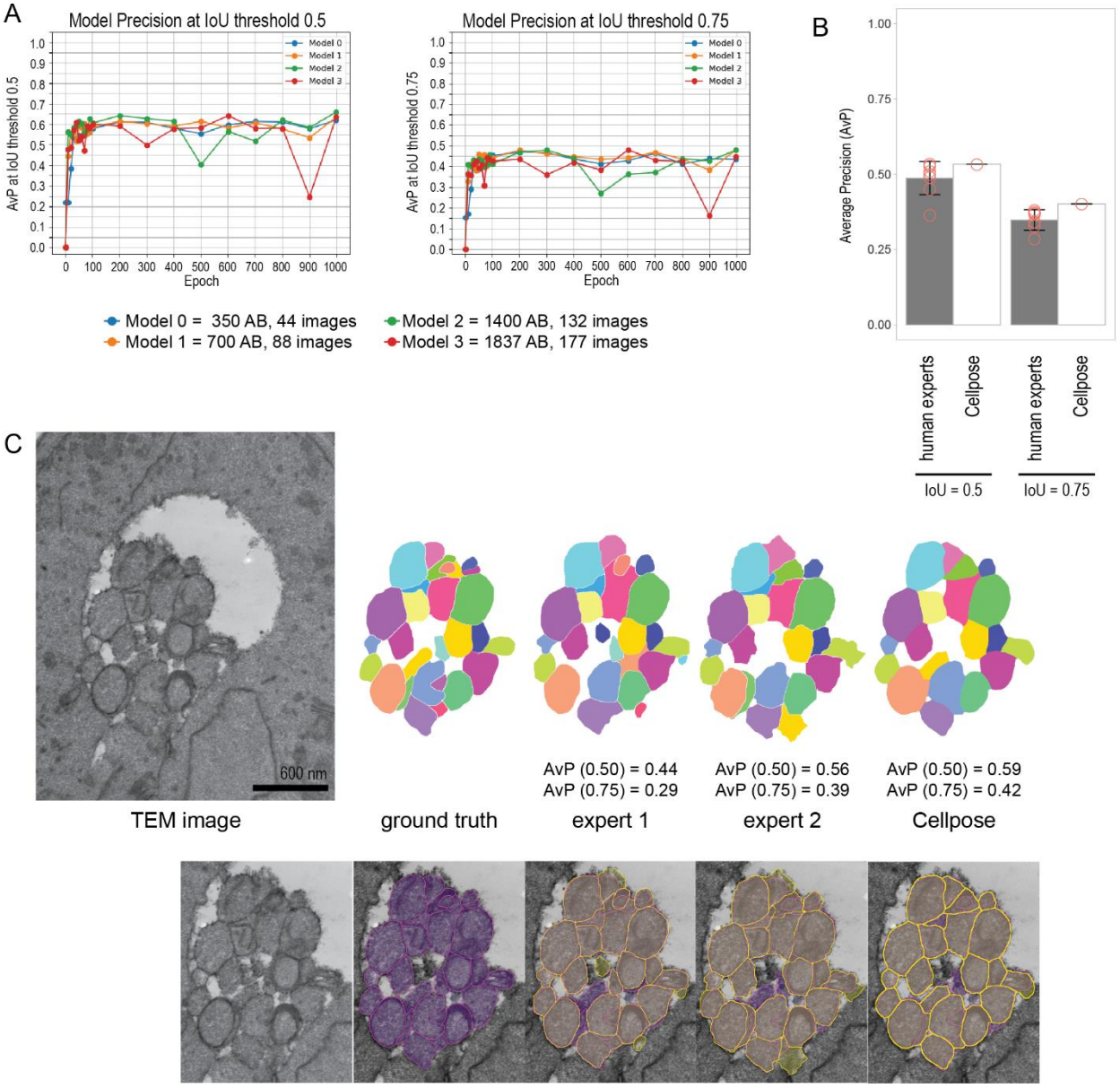

Figure 1. Evaluation of Cellpose 2.0 models trained to accurately predict ABs. (**A**) Performance of Cellpose models trained on increasing numbers of ABs for various epochs when tested on 51 images not in the training set, as measured by the average precision (AvP) at intersection over union (IoU) thresholds of 0.5 and 0.75. No additional improvement in performance was seen after 200 epochs of training, and very little improvement was seen beyond 350 ABs, showing that Cellpose 2.0 can be accurately trained on relatively little data. (**B**) Comparison of the best Cellpose model (Model 2, 201 epochs) to the AB segmentation performance of eight human

experts on nine challenging TEM images. All segmentations were compared to the "ground truth" segmentation of a ninth human expert who had overseen the ground truth segmentation of the training images, and AvP values were calculated at IoU thresholds of 0.5 and 0.75. Hollow circles represent the performance of individual segmenters, while bar graphs are the average performance and error bars indicate the standard deviation. Cellpose performed as well as any of the human experts on this task. (**C**) A portion of one of the nine TEM images and examples of various segmentations of that image, including the ground truth reference segmentation, segmentations by two separate human experts, and the segmentation generated by Cellpose using Model 2. AvP values relative to the ground truth segmentation are shown. In lower panels, magenta outlines indicate the ground truth segmentation and yellow outlines indicate the test segmentation. Scale bars: 600nm.

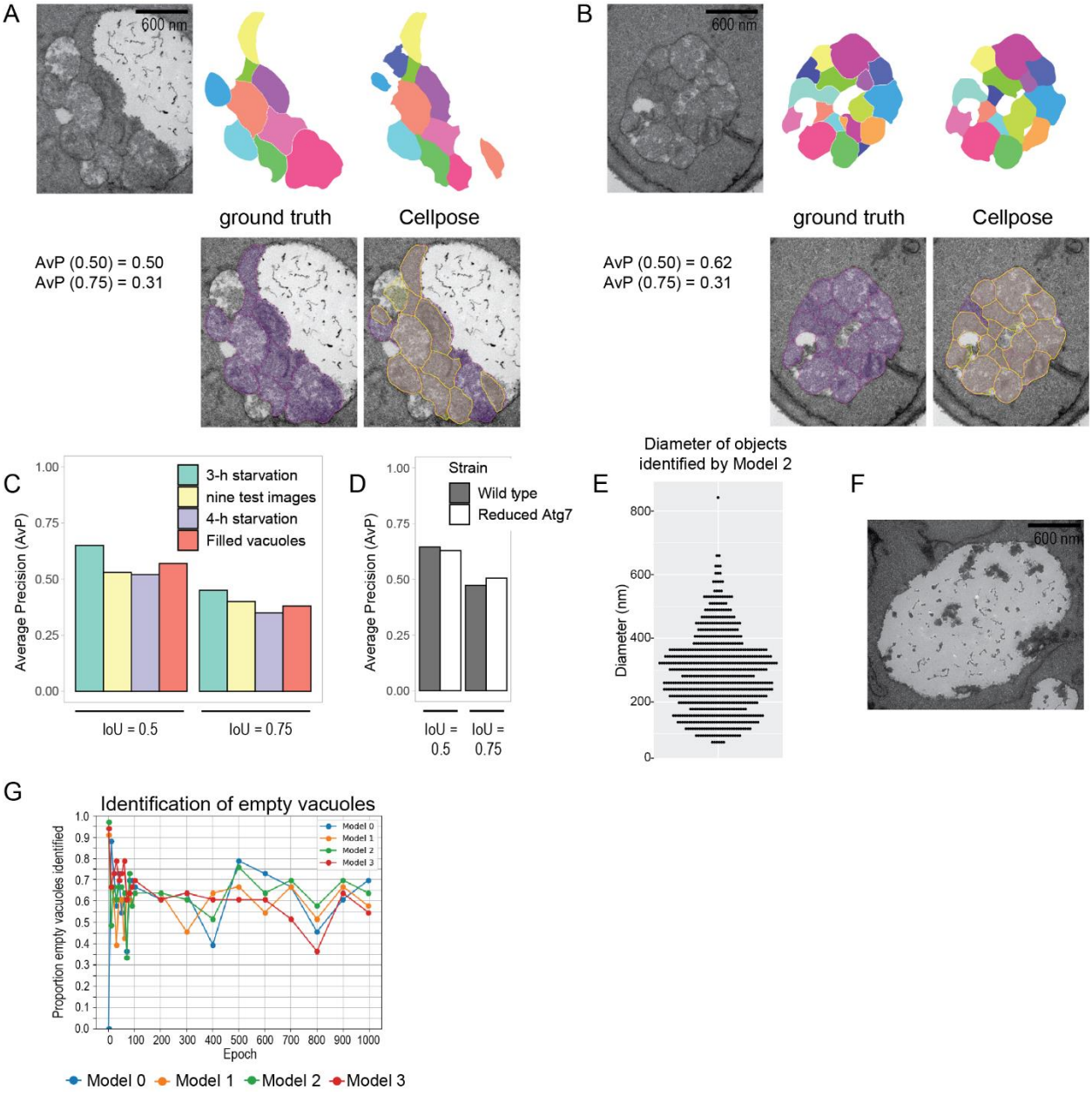
A
600 nm
ground truth
Cellpose
AvP (0.50) = 0.50
AvP (0.75) = 0.31
B
600 nm
ground truth
Cellpose
AvP (0.50) = 0.62
AvP (0.75) = 0.31
C
3-h starvation
nine test images
4-h starvation
Filled vacuoles
Average Precision (AvP)
IoU = 0.5
IoU = 0.75
D
Strain
Wild type
Reduced Atg7
Average Precision (AvP)
IoU = 0.5
IoU = 0.75
E
Diameter of objects identified by Model 2
Diameter (nm)
F
600 nm
G
Identification of empty vacuoles
Proportion empty vacuoles identified
Epoch
Model 0
Model 1
Model 2
Model 3

Figure 2. Macroautophagic bodies of varying sizes generated by different starvation times can still be recognized by Model 2. **(A**) Performance of Model 2 on ABs in the vacuole of a cell starved for four h instead of three. This was a typical result, with AvP values similar to the average of all forty 4-h starvation test images. In lower panels, magenta outlines indicate the ground truth segmentation and yellow outlines indicate the Cellpose segmentation. (**B**) Performance of Model 2 on an image selected because it showed a vacuole cross-section completely filled with AB cross-sections. This was a typical result, with AvP values similar to the average of all twenty images selected for this feature. (**C**) Comparison of the performance of Model 2 on different image test sets. The 3-h test set consisted of 51 images from the same samples as the training images, while the 4-h test set and the filled vacuoles test set were from different samples, and the nine test images were those used to compare the performance of Model 2 to that of human experts (Figure 1B). (**D**) Comparison of the performance of Model 2 on normal sized bodies (Wild type) and bodies from a mutant with reduced Atg7 that were an average of 20% smaller. Both types of bodies had been included in the training data. (**E**) Dot plot of the diameter of AB cross-sections labeled by Model 2 in the 3-h test set. The smallest cross-sections recognized were ~ 60 nm in diameter. N = 694 ROIs. (**F**) TEM image of the vacuole of an *atg9Δ pep4Δ* cell starved for 3 h. No macroautophagic bodies are present due to the lack of Atg9, but numerous smaller vesicles (multivesicular bodies and/or microautophagic bodies) can be seen due to the presence of functional ESCRT machinery in these cells. Model 2 did not identify any ROIs in this image. (**G**) The ability of Cellpose models to correctly identify empty vacuoles was tested on 33 images with no ABs; Cellpose's performance was relatively poor on this task and did not improve with training, likely because it was not possible to include empty vacuole images in the training set. Scale bars: 600nm.